工業技術研究院
Industrial Technology
Research Institute

採購規範書

1. 目的／用途說明：

為延續「提升山域事故救援效能五年中程計畫」之人工智慧語音整合應用成果，辦理系統移交、容器化服務交付與指定機房部署。以既有成果整理、必要介接調整、部署與技術移轉為主要範圍，使工研院得於指定環境內完成安裝、驗證、操作及後續移交維運。

2. 品名：

山域搜救智慧語音互動系統整合部署與資安維運服務。

3. 服務規格：

No	品項	需求說明
1	既有成果盤點與移轉規劃	• 盤點既有前後端功能、API、虛擬人素材、模型、知識庫、資料庫、設定、授權及外部依賴，列出沿用、調整與移轉項目。
2	前端 JavaScript 原始碼交付	• 交付本案範圍內完整且可編輯之前端 JavaScript 原始碼；採 TypeScript 或框架者，應一併提供原始專案、HTML、CSS、素材、依賴版本、建置指令及環境設定範例。
3	LLM 與 RAG 容器服務	• 提供 RAG 文件解析、切分、向量化、檢索及向量資料庫所需服務；因應知識文件新增、修訂或更新，重新執行文件處理、向量化及索引建置，更新 RAG 知識庫，使檢索與問答服務使用最新資料。
5	Docker 交付與服務管理	• 提供 ASR 、TTS、LLM 推論對話所需之後端所有必要映像檔之可匯入封裝檔，或可由工研院控制之映像倉庫副本，列明版本標籤、映像摘要及檔案校驗值。 • 提供 Docker Compose 或等效啟動設定、環境變數範本、磁碟掛載、連接埠、網路、GPU 指派、健康檢查、日誌及重啟設定。

No	品項	需求說明
6	指定臺灣機房部署支援	• 完成指定機房之環境檢核、安裝、容器匯入、模型載入、知識庫移轉、前後端串接、功能驗證及上線支援；依機房規定提供到場安裝及測試服務。 • 提出 CPU、RAM、GPU 型號／數量／顯示記憶體、儲存空間、作業系統、驅動及容器執行環境之最低與建議配置，並於指定硬體實測。 • 配合 IP、DNS、憑證、防火牆、連接埠及存取權限設定；無對外網路時，提供離線匯入套件及必要依賴，完成不連線之外部服務依賴檢查。
7	整合驗證與效能測試	• 驗證文字／語音輸入、ASR 轉錄、RAG 檢索、LLM 生成、TTS 播放、虛擬人互動及文件生成之完整流程。 • 整體系統與語音合成服務應支援至少 50 人同時在線使用；測試應明列在線人數、各類請求併發數、請求間隔、排隊方式及硬體配置。 • 提供延遲、成功率及資源使用紀錄，完成容器重啟、備份還原與異常處理測試，並提交問題改善及複測結果。
8	技術移轉與維運服務	• 提供部署與維運手冊、API 文件、備份還原程序、故障排除說明及教育訓練，協助委託方指定人員依文件自行啟停與復原服務。 • 驗收後提供 1 年保固、問題排除及本案交付範圍內之必要修正；交付修正版映像檔、前端程式及相應文件。
9	API 存取權杖與維運環境資安機制	• 整合既有 API Gateway 與身分驗證機制，於受保護 API 採用 Bearer Token 或相容之存取權杖驗證；依使用者、角色或服務帳號授予必要權限。 • 配合於指定機房設定管理與模型服務之網路存取限制、最小權限、必要連接埠、來源限制與呼叫頻率限制；使用 Cookie 驗證者，應設定適當安全屬性及跨站請求偽造防護。後端服務不得因直接連接埠存取而繞過驗證與授權。 • 將憑證更新、映像檔弱點修正與資安異常處理納入保固維運。

部署權責：工研院提供或協調指定機房、伺服器／虛擬機、GPU、作業系統使用權、網路、儲存及必要管理權限；廠商負責本案服務之容量規劃、相容性確認、安裝設定、移轉及測試。硬體採購、機房租用、線路與電費不列入本案基本服務費。

服務運作位置：正式運作之 LLM、ASR、TTS、RAG、資料庫均應位於指定機房；正式服務不得依賴未經核准之外部推論 API。

4. 執行期間：

自簽約日起 30 日曆天內完成；第一期規劃報告於簽約後 10 個工作天內提交，第二期系統交付與部署測試成果於簽約後 20 日曆天內提交，第三期整合驗證與技術移轉成果於簽約後 30 日曆天內提交。

5. 驗收標準：

（一）第一期規劃報告：提交既有成果與交付資產盤點、指定機房算力需求、部署計畫及工作時程，經委託方審查確認。

（二）第二期系統交付與部署測試成果：完成前端原始碼及後端映像檔交付，於指定機房完成部署與基本功能驗證，並提交部署紀錄、版本清單及初步測試報告。委託方指定人員應可依交付文件重建前端、匯入映像檔並啟動服務。

（三）第三期整合驗證與技術移轉成果：完成下列驗收項目；驗收發現之缺失應完成改善與複測，經委託方確認後辦理結案。

No	驗收項目	通過條件
1	前端與既有功能	原始碼可重新安裝依賴、建置及部署；文字與語音查詢、虛擬人至少 2 種動態反應、偏好設定記憶及 TTS 語速調整可正常操作。
2	AI 與資料移轉	LLM、ASR、TTS 及 RAG 於指定機房可呼叫；知識檢索案例均完成測試，核對來源、轉錄及語音播放結果。
3	交付與可復原性	依清單核對原始碼、映像檔；完成一次容器停止／重啟及一次知識庫與設定備份還原，復原後功能正常。
4	50 人在線與效能	至少 50 個有效使用者工作階段持續 30 分鐘，各階段依測試腳本執行問答及 TTS 播放；記錄 ASR、LLM、TTS 實際併發量、成功率及延遲，不得有服務崩潰或資料遺失。
5	機房與資料安全	確認服務位於指定機房、必要連接埠及權限已設定；無未授權對外傳輸、預設管理密碼或前端金鑰；離線情境依核定部署條件實測。
7	權杖驗證與權限控管	以有效、缺少、過期、遭竄改、已撤銷及權限不足之權杖分別測試；僅符合核定存取政策之請求得取得受保護資料。確認權杖更新、登出／停用失效及直接存取後端之管制符合設計。

6. 付款標準：

（一）契約總金額之100%：廠商於驗收後支付

- 提交第一期規劃報告含既有成果與交付資產盤點、指定機房算力需求、部署計畫及工作時程，經委託方審查確認，依契約辦理請款。
- 提交第二期成果報告含前端原始碼及後端映像檔交付，於指定機房完成部署與基本功能驗證，並提交部署紀錄、版本清單及初步測試報告，驗收通過後，依契約辦理請款。
- 提交第三期成果報告，完成指定機房整合驗證、教育訓練及全部交付，經最終驗收通過後，依契約辦理請款。

7. 訓練：■有　□無

廠商應提供至少 1 場教育訓練，地點或遠端或書面方式由委託方指定，內容包含前端建置、容器及模型啟停、知識庫更新、權杖與權限管理、憑證更新及常見故障處理。應交付電子教材與操作範例，並安排委託方指定人員實作演練。

8. 保固：■有　□無，由最終驗收完成日起算 1 年

保固包含交付程式、映像檔、設定與介接之瑕疵修正、部署問題排除，以及影響本案運作之必要相容性與安全修正。廠商應交付修正版、版本異動說明及更新／回復程序，並配合指定機房完成驗證。新增功能、模型重新訓練、大量新增資料處理或更換硬體所衍生之工作，應經雙方確認範圍及費用後辦理。

9. 服務：□有　■無

10. 廠商交付文件：■有　□無

各期報告電子檔各 1 份，提供可編輯格式及 PDF；技術資產應以委託方可下載、保存與操作之形式點交，至少包括：

（一）第一期規劃報告：成果盤點、架構與部署規劃、最低與建議算力與時程。

（二）第二期成果報告：前端完整原始專案與建置成果、後端 Docker 映像檔、部署設定、權杖及權限設定說明、機房部署及基本測試紀錄。

（三）第三期成果報告：整合及效能測試報告、缺失修正與複測紀錄、權杖及維運環境資安驗證紀錄、、教育訓練教材與紀錄、操作維運手冊、完整交付清冊與保固服務資料。

11. 其他說明事項：

（一）禁止轉包或分包：得標廠商不得將本案履行項目之全部或主要部分交由他人代為履行；相關違約處理與損害賠償依契約約定辦理。

（二）保密責任：廠商及參與人員因履約知悉或取得之文件、語音、對話、知識庫、帳號及機房資訊，未經委託方事前書面同意，不得洩漏、交付第三人或用於本案以外之訓練及用途。履約結束時應完成資料返還及廠商端副本清理，依法或依契約應留存者除外。

（三）使用與維護權利：本案交付之前端原始碼及客製介接成果，應使委託方及其指定維運單位得為本案目的使用、複製、修改、建置、備份及部署；具體權利歸屬依契約約定。後端以可運作之 Docker 映像檔及必要模型、設定與維運資料交付；既有通用後端框架之原始碼不列為當然交付項目，但不得影響約定功能及維運。

（四）第三方及模型授權：廠商應確認程式套件、模型權重、聲音及虛擬人素材之授權適用於本案部署與使用，列明再散布、修改、商用、使用期間與數量限制。不得以交付映像檔取代必要授權；需由委託方申請者應於報價時揭露並協助完成，持續使用費及保固期後費用應分項列明，不得留有未揭露之必要訂閱或啟用條件。